Genetic Programming for Evolutionary Feature Construction

Hengzhe Zhang

Supervisor: Mengjie Zhang, Bing Xue, Qi Chen

Victoria University of Wellington

08/08/2024

Table of Contents

1 Introduction

2 Fundamental Research

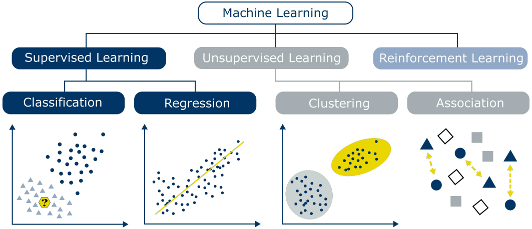
3 Applied Research

4 Conclusion

Introduction

Machine Learning

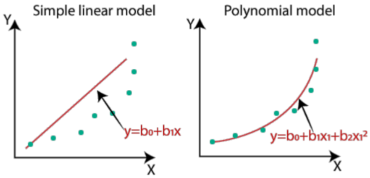
- A computational method that uses past experiences to generate accurate predictions for future data.
- Applications: intelligent agriculture, disease diagnosis, and natural disaster prediction, etc.



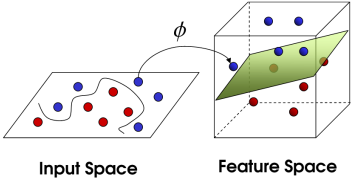
Machine Learning

Feature Construction

- The general idea of feature construction is to construct a set of new features $\{\phi_1, \ldots, \phi_m\}$ to enhance the learning performance of machine learning algorithms on a given dataset $\{\{x_1, y_1\}, \ldots, \{x_n, y_n\}\}$ compared to learning on the original features $\{x^1, \ldots, x^p\}$.



(a) Feature Construction on Regression



(b) Feature Construction on Classification

An Example

- Body Mass Index (BMI) is an example of a feature constructed by combining weight and height measurements.
- BMI is associated with obesity levels in patients from New Zealand.

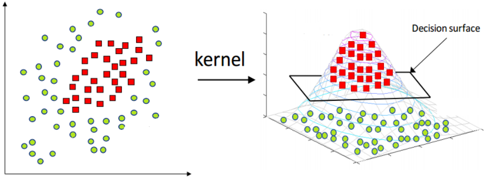
$$\text{BMI} = \frac{\text{Weight (in kilograms)}}{\text{Height}^2 \text{ (in meters)}}$$

Body Mass Index

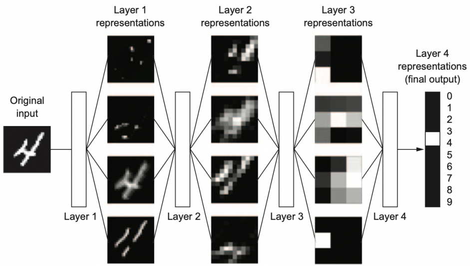
[1] Mohammed A Moharram et al. (2020). **"Correlation between epicardial adipose tissue and body mass index in New Zealand ethnic populations".** In: *The New Zealand Medical Journal (Online)* 133.1516, pp. 22–5.

Feature Construction Methods

- Kernel Methods
- Deep Learning



(a) Kernel Methods



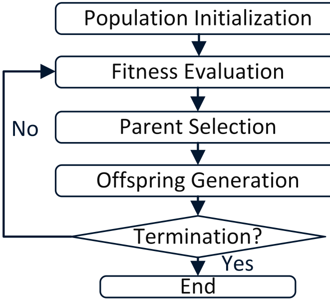
(b) Deep Learning

Genetic Programming

- Genetic programming is a population-based search method with variable-length representation that aims to search for a computer program capable of solving a given task.

Advantages:

- Flexible Length of Representation
- Population-Based Search Algorithm
- Gradient-Free Search Mechanism
- Good Interpretability

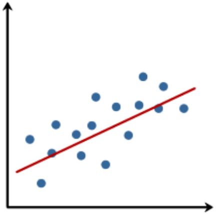


The evolution process of GP.

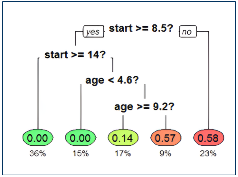
Fundamental Research

Feature Construction for Ensemble Learning:

- For Linear Models (CIM 2023)[1]
- For Decision Trees (TEVC 2021)[2]



(a) Linear Model



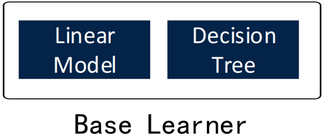
(b) Decision Tree

[1] Hengzhe Zhang, Qi Chen, Bing Xue, Banzhaf Wolfgang, et al. (2023). **"MAP-Elites for Genetic Programming-based Ensemble Learning: An Interactive Approach"**. In: *IEEE Comput. Intell. Mag.*

[2] Hengzhe Zhang, Aimin Zhou, and Hu Zhang (2022). **"An Evolutionary Forest for Regression"**. In: *IEEE Trans. Evol. Comput.* 26.4, pp. 735–749.

Feature Construction for Heterogeneous Ensembles:

- Ensemble of linear models and decision trees (TEVC 2023)
- Outperforms state-of-the-art methods such as Random Forest, XGBoost, and LightGBM

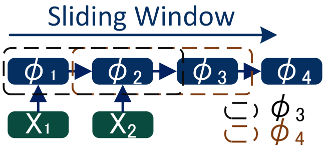


Linear Models and Decision Trees

[1] Hengzhe Zhang, Aimin Zhou, Qi Chen, et al. (2023). **"SR-Forest: A Genetic Programming based Heterogeneous Ensemble Learning Method"**. In: *IEEE Trans. Evol. Comput.*

Modular Feature Construction:

- **Modularity:** Outputs from one GP can be inputs for other GP trees.
- **Sliding Window:** Provides more flexible context compared to layer-by-layer modularity.
- **Performance:** Outperforms state-of-the-art methods like PS-Tree in accuracy and model size.

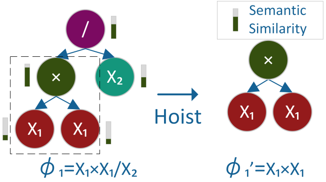


Modular Feature Representation

[2] Hengzhe Zhang, Qi Chen, Bing Xue, Wolfgang Banzhaf, et al. (2023b). **"Modular multi-tree genetic programming for evolutionary feature construction for regression"**. In: *IEEE Transactions on Evolutionary Computation*.

Semantic Hoist Mutation:

- **Semantic Mutation:** Moves the most informative subtree to the root based on target semantics.
- **Theoretical Guarantee:** No increase in Vapnik–Chervonenkis dimension.
- **Performance:** Outperforms state-of-the-art methods like PS-Tree in accuracy and model size.



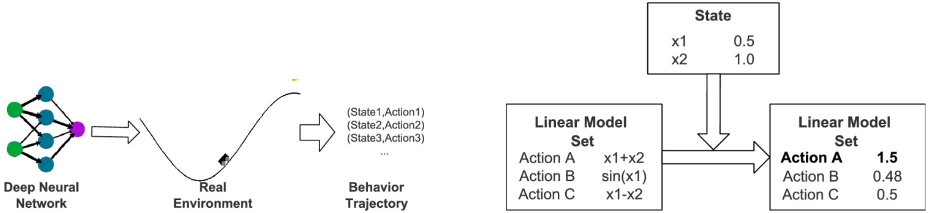
Semantic Hoist Mutation

---

[3] Hengzhe Zhang, Qi Chen, Bing Xue, Wolfgang Banzhaf, et al. (2023a). **"A semantic-based hoist mutation operator for evolutionary feature construction in regression"**. In: *IEEE Transactions on Evolutionary Computation*.

Feature Construction for Reinforcement Learning:

- Applied to reinforcement and imitation learning (Comp. Intell. Syst. 2020)[1]
- Enables linear models to achieve performance comparable to deep neural networks



(a) Data Collection **(b)** Imitation Learning

[1] Hengzhe Zhang, Aimin Zhou, and Xin Lin (2020). "**Interpretable policy derivation for reinforcement learning based on evolutionary feature synthesis**". In: *Comp. Intell. Syst.* 6, pp. 741–753.

Applied Research

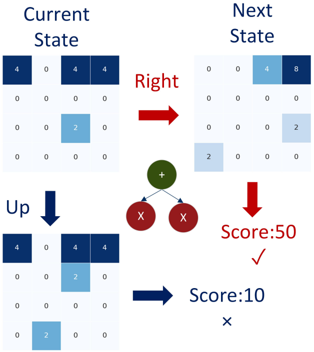
Objective: Use GP for interpretable control in the 2048 game.

Method:

- **GP Strategy:** Evolve scoring functions with domain features like Monotonicity.
- **Planning:** Multi-step lookahead for action evaluation.

Results:

- Model: $neg(X_2 \times 29.83)$, where X_2 is monotonicity.
- Score: 19767 (Found 2048!).
- Winner of GECCO 2024 Competition.



GP for Playing 2048

Conclusion

Conclusion

- Evolutionary feature construction is an effective technique for improving learning performance.

Future Directions

- Applying feature construction to real-world applications, including but not limited to cyber-marine seafood, climate change, and marine genomics.

Thanks for listening!

Email: Hengzhe.zhang@ecs.vuw.ac.nz

GitHub Project: https://github.com/hengzhe-zhang/EvolutionaryForest